This readme file is for "Anomalies as New Hedge Fund Factors" by Yong Chen, Sophia Zhengzi Li, Yushan Tang, and Guofu Zhou published in *JFQA*.

Software Package

Python is used for coding, with the code tested under Python 3.9.19. Below are the Python packages along with their corresponding versions necessary for replicating the results. Please note that results may vary slightly if you are using a different version. Packages and Versions:

- numpy 1.24.3
- pandas 1.5.3
- sklearn 1.0.1
- statsmodels 0.13.5
- scipy 1.10.1
- matplotlib 3.7.1
- seaborn 0.12.2
- linearmodels 6.1
- lifelines 0.27.4

Master Code

The `Master Code' produces and organizes the results based on input datasets and files. **To replicate the process of generating all result figures and tables in the main manuscript, simply configure the working environment and execute the `Master Code'.** Please note that the pseudo data sets are provided solely to demonstrate the required file formats and to ensure that all programs execute successfully. They do not reproduce the key results, as the original data are subject to copyright.

- **`MASTER_CODE.py'** utilizes input datasets from the `PseudoData' folder, produces the following figures and tables which are subsequently stored in `Figure' and `Table' folders:
    - Figure 1: Variance explained ratios across principal components
    - Figure 2: Distribution of hedge fund alphas
    - Figure 3: Distribution of hedge fund alphas: BarclayHedge
    - Figure 4: Distributions of bootstrap t-statistics

    - Table 1: Descriptive statistics for hedge fund returns
    - Table 2: Descriptive statistics for factors
    - Table 3: Factor selection via adaptive LASSO
    - Table 4: Nested model selection
    - Table 5: Model comparison
    - Table 6: Model comparison: Fund alpha distribution
    - Table 7: Model comparison: BarclayHedge

- Table 8: Model comparison: Out-of-sample period
- Table 9: Model comparison: Arbitrage trading
- Table 10: Model comparison: Alternative models
- Table 11: Flow-performance relation
- Table 12: Fund failure likelihood
- Table 13: Bootstrap analysis of fund alphas
- Table 14: Fund alpha and characteristics

Pseudo Data Sets

The required input pseudo datasets and files are organized as follows:

- **`FUNDID_INCEPT.dta'** contains fund masked IDs and inception dates.
- **`MONTH_ADDITIONAL_FACTOR.xlsx'** contains monthly time series of four additional factors.
- **`MONTH_FACTOR.xlsx'** contains monthly time series of risk factors.
- **`MONTH_FACTOR_201909_202112.xlsx'** contains monthly time series of risk factors used in out-of-sample analysis.
- **`MONTH_FF5F.xlsx'** contains monthly time series of Fama-French five factors.
- **` NAT_RET.dta'** contains monthly returns, NAT scores, and market equity values for individual stocks.
- **`PSEUDODATA1.pkl'** contains pseudo data of hedge fund excess returns and characteristics from TASS and HFR database.
- **`PSEUDODATA2.xlsx'** contains pseudo data of hedge fund index returns from BarclayHedge database.
- **`PSEUDODATA3.xlsx'** contains fund masked IDs, styles, and inception dates from BarlceyHedge database.
- **`PSEUDODATA4.xlsx'** contains pseudo data of hedge fund excess returns from BarlceyHedge database.
- **`PSEUDODATA5.csv'** contains pseudo data of hedge fund portfolio excess returns from TASS database for out-of-sample analysis.